Research in Astronomy and Astrophysics, 24:055013 (18pp), 2024 May https://doi.org/10.1088/1674-4527/ad3c6e
© 2024. National Astronomical Observatories, CAS and IOP Publishing Ltd. Printed in China and the U.K.

Real-time Abnormal Detection of GWAC Light Curve based on Wavelet Transform Combined with GRU-Attention

Hao Li[1], Qing Zhao[1], Long Shao[1], Tao Liu[1], Chenzhou Cui[2], and Yunfei Xu[2]
[1] College of Artifcial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China; 22835030@mail.tust.edu.cn
[2] National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China
Received 2024 January 4; revised 2024 March 10; accepted 2024 March 26; published 2024 May 10

Abstract

Nowadays, astronomy has entered the era of Time-Domain Astronomy, and the study of the time-varying light curves of various types of objects is of great significance in revealing the physical properties and evolutionary history of celestial bodies. The Ground-based Wide Angle Cameras telescope, on which this paper is based, has observed more than 10 million light curves, and the detection of anomalies in the light curves can be used to rapidly detect transient rare phenomena such as microgravity lensing events from the massive data. However, the traditional statistically based anomaly detection methods cannot realize the fast processing of massive data. In this paper, we propose a Discrete Wavelet (DW)-Gate Recurrent Unit-Attention (GRU-Attention) light curve warning model. Wavelet transform has good effect on data noise reduction processing and feature extraction, which can provide richer and more stable input features for a neural network, and the neural network can provide more flexible and powerful output model for wavelet transform. Comparison experiments show an average improvement of 61% compared to the previous pure long-short-term memory unit (LSTM) model, and an average improvement of 53.5% compared to the previous GRU model. The efficiency and accuracy of anomaly detection in previous paper work are not good enough, the method proposed in this paper possesses higher efficiency and accuracy, which incorporates the Attention mechanism to find out the key parts of the light curve that determine the anomalies. These parts are assigned higher weights, and in the actual anomaly detection, the star is detected with 83.35% anomalies on average, and the DW-GRU-Attention model is compared with the DW-LSTM model, and the detection result f1 is improved by 5.75% on average, while having less training time, thus providing valuable information and guidance for astronomical observation and research.

Key words: methods: data analysis – stars: variables: general – techniques: photometric

1. Introduction

Early warning of light curve anomalies is one of the important research directions in astronomy. In the field of astronomy, light-variation curves are curves of the radiant luminosity of celestial objects plotted over time, which can provide important information about the physical properties of celestial objects and their evolutionary patterns. As the field of astronomy enters the era of time-domain astronomy, many international and domestic astronomical telescopes have utilized cutting-edge detection techniques, which have captured a large number of light-variation curves of various types of celestial objects in a large number of systematic ways, and anomalous detection techniques of light-variation curves can effectively detect non-periodic phenomena such as supernova outbursts, gamma-ray bursts, and microgravitational lensing events. however, because of the transient and rare nature of these phenomena, the data to be processed are more voluminous. Data to be processed is even more massive, it is difficult to adopt the traditional statistical-based anomaly detection method. Therefore, the study of intelligent algorithms based on deep learning is the trend of development. However, due to the influence of various factors, the light curves are often contaminated by different degrees of noise, and the analysis results are easily distorted, while the astronomical light curves are very different in the cycle law, and there are multi-scale features, and the direct use of deep time series prediction methods based on recurrent neural networks, long-short-term memory (LSTM), etc. is still ineffective. Therefore, it is of great significance to study the light curve anomaly warning method suitable for massive data processing and based on a new depth model for carrying out large-scale time-domain astronomical research.

Ground-based Wide Angle Cameras (GWAC) is a large-field-of-view, high temporal-resolution optical observing system led by the National Astronomical Observatories, Chinese Academy of Sciences (NAOC), which is mainly used for detecting and tracking Gamma-Ray Bursts (GRBs) and other transient objects. GWAC has an observing field of view of more than 2000 square degrees, with a depth of detection up to 16 mag, and a time resolution of 15 s, allowing real-time

monitoring of a large area of the sky and capturing astronomical phenomena such as extreme relativistic jets, neutron star-conjugated gravitational wave events, etc. The GWAC started its trial operation at the Xinglong Base of the National Astronomical Observatories in 2016, and has observed more than 10 million light curves so far.

Automation Algorithms Related Technologies has been widely used in the astronomical field. Van Doorsselaere et al. (2017) proposed an automated flare detection and characterization algorithm to analyze stellar flares observed by the Kepler mission. They developed an automated flare detection and characterization algorithm, that discovered flares from new candidate A-type stars and 653 giant stars, demonstrates that automated algorithms can be well applied to astronomical data processing. Vida & Roettenbacher (2018) explored the use of machine learning tools to identify and analyze flares in Kepler data, the RANSAC algorithm was used to detect anomalies, and machine learning methods were used to identify flare events in light curves, realising innovations in machine learning for astronomical anomaly detection. Breton et al. (2021) introduced a machine learning tool called ROOSTER for automatically determining the rotation period of stellar surfaces using Kepler light-variation curves, the introduction of ROOSTER provides new ideas for efficiently analyzing large stellar photometric data sets. Althukair & Tsiklauri (2023) wrote and used an automated flare detection Python script to search for super-flares on main-sequence stars of types A, F, G, K, and M in Keplers long-cadence data from Q0 to Q17, illustrating the higher efficiency of automated scripts for long-cadence data processing.

Traditionally, astronomical light curves have been analyzed for anomalies using statistical methods or traditional machine learning methods. Bi et al. (2018) proposed an enhanced Autoregressive, Regressive and Integrated Moving Average (ARIMA) model to analyze light curves from GWAC collection, and its experimental results demonstrated its usefulness for anomaly detection, we propose this as an improved ARIMA model. It has shown promising results for anomaly detection as the improved ARIMA model shows. Feng et al. (2017) proposed a time-series analysis model "DARIMA," which can identify the first anomaly of all light curves. Lu (2022) proposed some solutions for non-uniform sampling data of light curves, such as Discrete Fourier Transform (DFT) method, Discrete Correlation Function (DCF) method, Lomb-Scargle Periodogram (LSP) method, and Weighted Wavelet Z-transform (WWZ) methods, etc. Kalaee & Hasanzadeh (2019) conducted research into the periodic behavior and variability of R Scuti stars using power spectral density and Fast Fourier Transforms to examine light curves between 1970 and 2017 time series analysis. Deb & Singh (2009) conducted similar research using Fourier decomposition and principal component analysis for light variation curves to demonstrate peak-finding capabilities using Fourier transform analysis for astronomical light variation curves. Huang (2019) developed Random Forest-based screening methods for transient and variable sources. She extracted features of stars using Principal Component Analysis before classifying with Random Forest as the classifier, then performed observational data screening to demonstrate the viability of using machine learning methods for screening transient and variable sources. Yu et al. (2021) provided an overview of machine learning's role in the analysis of light-variation curves, which demonstrated its usefulness for peak-finding in astronomical light-variation curves in an age of big data. Machine and deep learning play an increasingly significant role in exploring light-variation curves amidst an ocean of big data. Traditional statistics and machine learning methods require substantial computational resources for long time series data, while still missing some key information or complex patterns present. Therefore, finding models with higher efficiency and accuracy would help address such problems more efficiently.

Researchers have developed innovative time series models using deep learning. These models use deep learning to extract patterns and characteristics from complex time-series data, to improve forecasting accuracy and efficiency. These advances not only show the power of deep-learning techniques for time series analyses, but they also represent a significant milestone in methods for processing sequence data and capturing temporal dependences, as well as forecasting future trends. Deep learning techniques for time series predictions typically rely upon recurrent neural networks (RNN), and their variants, such as the LSTM or gated recurrent units (GRU), that use hidden states to capture historical information and dynamic dependencies in time series.

Deep Learning for Astronomical Data Processing has seen some pioneering applications. For instance, Burhanudin et al. (2021) proposed an RNN neural network classifier to recognize incomplete light curves, while Lu et al. (2018) devised a DRNN deep neural network to optimize photometric variations prediction. Xu et al. (2018) examined the use of deep learning as an aid for processing astronomical big data and presented research results from Solar Key Laboratory of National Astronomical Observatories, Chinese Academy of Sciences to illustrate its application in this domain. Boone (2021) employed a deep learning model to generate transient light-variation curves, which demonstrated its excellent potential application in astronomical big data analysis. Regarding time series neural network models, time series neural networks proved effective at simulating transient light variation curves, Zhang & Zou (2018) implemented a light curve warning based on long short-term memory (LSTM) network, using long short-term memory network to detect anomalies in light curves, and Chakraborty (2019) conducted a warning effect test with RNN-LSTM data, demonstrating the LSTM algorithm's success at solving light curve anomaly detection problems, but struggling

to perform in long time series due to its complex network structure. They tested both its detection abilities as well as performance on longer time series using simulation data, showing it can perform adequately on short term series but poorly on longer ones due to more complex network structures. Yan et al. (2020) proposed a real-time anomalous light curve warning model utilizing a GRU network for real-time anomalous light curve detection and warning, using collected light curve data as training to train it to predict star brightness at any moment, when this mismatch exceeds an agreed-upon threshold value, an anomaly is recognized and warned upon. Experimental results demonstrate that, the traditional neural network model can successfully apply to astronomical light curve anomaly detection, however its prediction accuracy and prediction effect need further improvements as some dependencies present within light curve data remain unexamined and untested. This suggests that it has potential use as an anomaly warning system, but it still requires further exploration and training of its capabilities before being put to work effectively. In recent years, deep learning models based on attention mechanisms have held new promise for anomaly detection in large-scale light curves. Bowles et al. (2021) introduced the Attention model to solve the problem of classifying interpretable radio galaxies in astronomy, evaluating cyclic isotropy and dihedral isotropy of various orders and showing that isotropy is included as a priori. Both the reduction in the number of training sessions required to fit the data and the improvement in performance amply demonstrate the value of the Attention model for applications in astronomy. However, due to the observation conditions, instrumental noise, atmospheric effects, and other factors, the light-variation curves are often contaminated to different degrees, resulting in lower signal-to-noise ratios and distorted analysis results. Therefore, noise reduction of light-variation curves is one of the important steps in astronomical data analysis.

Xu et al. (2022) proposed a post-training quantization preprocessing method for convolutional neural network models based on outlier removal, that can effectively reduce quantization error while increasing accuracy and robustness of quantization models. This shows that outlier removal is feasible on time series training but should not be limited to isolated data, to improve outlier removal, the time window outlier removal method was utilized instead thereby eliminating instances of mistaken removals.

Wavelet noise reduction is a popular tool for both noise reduction and feature extraction, employing wavelet transform to dissect signals into wavelet coefficients of different scales and frequencies, then filter or compress these coefficients according to different thresholding rules in order to identify noise components, before reconstructing a reduced signal by inversive wavelet transform. Wavelet noise reduction provides excellent time-frequency localization capabilities, as well as accommodating for non-stationary and multi-resolution features of signals. Sasal et al. (2022) have extensively documented these benefits in their work, for instance. W-Transformer is an univariate time series representation learning framework for wavelet-based transformer encoder architecture, which using MODWT decomposition of time series data, and building local transformers to accurately capture any nonsmoothness or long-range nonlinear dependencies within time series data. Ma et al. (2022) combined wavelet transform with neural network and proposed an automatic search method for X-ray astronomical burst events based on wavelet transform and convolutional neural network. The results obtained proved the effectiveness of this method in the problem of finding peaks in optical variables. Wavelet transform is a very effective mathematical tool in early warning of light curve anomalies. The core role of wavelet transform is to analyze signals in the time domain and frequency domain. Unlike the traditional Fourier transform, the wavelet transform provides time and frequency information, which makes it particularly suitable for analyzing non-stationary signals, such as astronomical light curves, whose properties may change over time. When performing wavelet transformation on the light curve, the low-frequency part mainly describes the slow trend of the signal and represents the long-term trend of the celestial body. The high-frequency portion captures the rapid changes and details of the signal, which often contains noise or sudden events caused by instrument errors or short-term anomalous celestial phenomena. The implementation of this method shows that neural networks can provide a more effective processing model for wavelet transform results.

In summary, the time series neural network is an excellent time series detection model, in which the GRU model has higher efficiency and adaptability in the processing of long time series. The attention mechanism is easier to capture the dependency relationship within the time series, and can give higher weight to the key part of the decision of anomaly. Wavelet transform has a good effect on data processing, which can make the masked features in the light curve be mined. Therefore, this paper proposes a GWAC light curve anomaly early warning model based on the combination of wavelet transform and GRU-Attention. The excellence of combining wavelet transform with time series neural network and Attention mechanism lies in that they can complement and enhance each other, the discrete wavelet transform is introduced for data enhancement and frequency domain information acquisition. Because the discrete wavelet transform has translation invariance and variable resolution characteristics compared with methods, such as the Fourier transform, it is expected to better solve the problem of multi-resolution, waveform multi-scale information extraction of astronomical time series data, the Attention allows the model to focus more on the parts, that are decisive for anomaly monitoring, the wavelet transform can provide richer and more stable input features for subsequent training of neural networks, and neural

Table 1
Light Curve Data Field Information Selected from GWAC

Columns	Meaning	Example
StarId	Unique identifier	ref_033_16810765-G0013_482792_15702
Type	Rare event types	flare star
Path	Path	AstroSet/033_16810765-G0013
R.A.	Right ascension	176.768005
decl.	decl.	70.031097
Length	Curve length	4041
JD	Observation time	2458508.1344330
Magnorm	Magnitude	9.11
Mage	Magnitude error value	0.06

networks can provide more flexible and powerful output models for wavelet transform.

2. Data

2.1. Light Curve Data

As shown in Table 1, the light curve data used in this paper are from the Tianchi Astronomical Time Domain Dataset,[3] collected by the GWAC Astronomical Survey Facility. The data set has a total of 766,576 light curves that have been calibrated with relative fluxes, with an observational time span of 6 months, and a temporal sampling rate of 1 data point per 15 s for the continuous portion of the light curves, with a total of 26 observational sky regions. The data set has been labeled with stellar types and includes information on 18 short-lived rare-object light-variation events.

2.2. Sliding Window Method for Outlier Removal

The sliding window method is a method for outlier detection and removal, which has the advantage of effectively eliminating noise and outliers from data, thus enhancing the reliability and robustness of data analysis. The core principle of the sliding window method is that for each data point, a window of fixed length is selected as its center, and then the corresponding statistics, such as mean, variance, median, etc., are calculated based on the data within the window, and based on the comparison of these statistics with the preset thresholds or standard deviations, it is determined whether the data point is an outlier, and if it is, it is rejected. The observed data of light curves are usually affected by many factors, such as atmospheric refraction, instrument error, occlusion, missing, etc. These factors will lead to the existence of outliers in the data, which will affect the analytical tasks such as feature extraction, classification, regression, etc. of the light curves, and reduce the accuracy and efficiency of the automated analysis of the stars. Therefore, using the sliding window method to preprocess the light-variation curve data can significantly improve the quality of the data, and thus enhance the performance of the anomaly warning for stars. The formula for the outlier determination rule is:

$$\begin{cases} > \frac{1}{N}\sum_{i=1}^{N} \text{windows}_i + \text{threshold} \\ * \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\text{windows}_k - \frac{1}{N}\sum_{i=1}^{N}\text{windows}_i\right)^2} \\ < \frac{1}{N}\sum_{i=1}^{N} \text{windows}_i - \text{threshold} \\ * \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\text{windows}_k - \frac{1}{N}\sum_{i=1}^{N}\text{windows}_i\right)^2} \end{cases} \quad (1)$$

where denotes the data contained in the sliding, N denotes the size of the sliding window, and *threshold* denotes the customized threshold.

The sliding window effectively removes the noise from the original light curve while maintaining the anomalous features of the original light curve. Here, the anomalous objects with StarId of ref_022_15730595-G0013_391462_6330 and ref_044_16280425-G0013_364820_9174 are culled for the outlier, and the light curves before and after completing the culling are shown in the lower part of Figure 1.

The standard deviation before and after completing the culling is shown in Table 2, and it can be seen that the trend of the curves after culling the outliers is more stable, while retaining the basic features of the light-variable curves, which is convenient for the training of neural networks.

2.3. Data Normalization

Each light-variation curve data in this data set has a unique identification StarId, the StarIds of the two stars selected in this paper are ref_033_16810765-G0013_482792_15702 and ref_044_16280425-G0013_364820_9174 respectively. the difference between different light-variation curves is very large, which greatly increases the time for model training and the difficulty of convergence, in order to eliminate the effect to some extent, the data is normalized here, and the normalized data is beneficial for accelerating the convergence of gradient descent, because it ensures that all the features are at the same scale and reduces the training time. When the input features are on the same scale, parameter initialization is more efficient, which is conducive to the stability and performance of model training. It also ensures that all features are involved in the model learning process with the same importance, which improves the effectiveness of weight updating.

This paper uses Min–Max Normalization, a process that scales all data points to between 0 and 1, maintaining the original distribution and proportions in the data. The formula

[3] 2021, Tianchi. AstroSet Dataset, [Onlined], https://tianchi.aliyun.com/dataset/88856/ 5, 20, 21

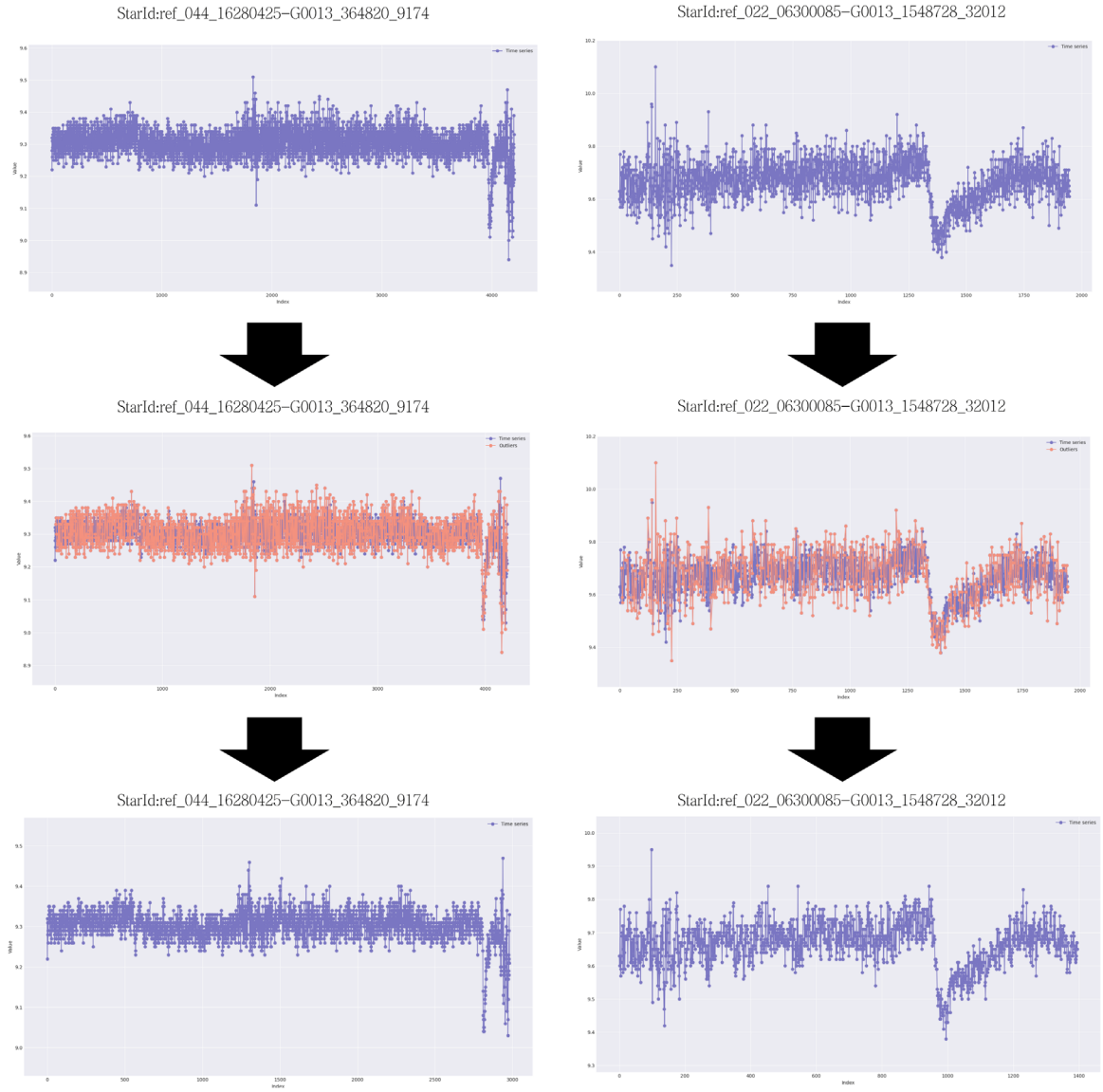

Figure 1. Comparison of light curve outliers before and after removal.

Table 2
Standard Deviation before and after Removal of Outliers from the Light Curve

Unique Identifier	Standard Deviation before Removal	Standard Deviation after Removal
ref_022_15730595-G0013_391462_6330	0.0803	0.0678
ref_044_16280425-G0013_364820_9174	0.0451	0.0374

is:

$$S = (x - x_{\min})/(x_{\max} - x_{\min}) \quad (2)$$

where x is the value of the original data, $x_{\min}$ is the minimum value of the sample data, which represents the minimum value in the original data, $x_{\max}$ is the maximum value of the sample data, which represents the maximum value in the original data.

We here normalize two stars with StarIds of ref_033_16810765-G0013_482792_15702 and ref_044_16280425-G0013_364820_9174, respectively, and the time-series images before and after the normalization are shown in Figure 2, which shows that the data are

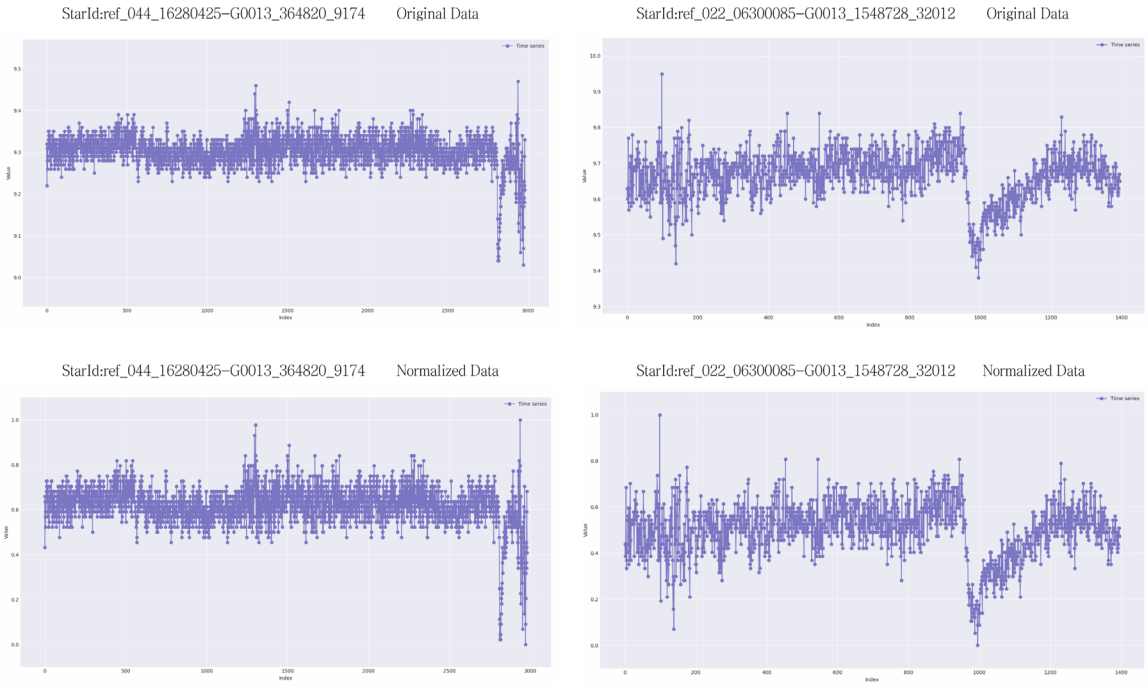

Figure 2. Light curve before and after normalization.

compressed to the same scale while retaining the features that the time-series data originally had.

3. Methodology

3.1. Feature Extraction and Signal Denoising based on Wavelet Transform

Wavelet transform is a mathematical tool widely used in signal processing and image processing, which can decompose a signal or image into wavelet coefficients of different scales and positions, thus realizing multi-resolution analysis. Wavelet transform has gone through many stages, from the beginning of continuous wavelet transform gradually developed to discrete wavelet transform, and then to wavelet packet transform and multidimensional wavelet transform, and so on. The theory and application of wavelet transform have also been deepened and expanded, involving signal denoising, image compression, image fusion, pattern recognition, feature extraction, and other fields.

Wavelet transform requires the selection of appropriate wavelet basis functions and decomposition layers to do the discrete wavelet transform (DWT) on the light-change curve to obtain wavelet coefficients at different scales. The selection of the wavelet basis function can be determined according to the characteristics and objectives of the signal, and generally requires good orthogonality and tight support. The selection of the number of decomposition layers can be determined according to the length of the signal and the distribution of the noise, which generally requires that the main information of the signal can be concentrated in the low-frequency part, while the main energy of the noise is dispersed in the high-frequency part. In this study, we choose sym8 wavelet as the wavelet basis function, which is a kind of symmetric wavelet with 8th order vanishing moments, which can fit the smoothness and abruptness of the light change curve better. We choose 6-layer decomposition, so that the light-change curve can be decomposed into one approximation coefficient and six detail coefficients, which correspond to different frequency ranges, and the decomposition principle is shown in Figure 3, CD is the detail coefficient, which is a high-frequency signal and is obtained by a high-pass filter, and CA is the approximation coefficient, which is a low-frequency signal and is obtained by a low-pass filter. The two stars with StarIds of ref_033_16810765-G0013_482792_15702 and ref_044_16280425-G0013_364820_9174, respectively, are decomposed to obtain the low-frequency part and the high-frequency part shown in Figure 4. The low-frequency part mainly describes the trend of the signal's slow change in the light curve which represents the long-term trend of the object. The high-frequency part captures the fast changes and details of the signal, which often contain noise or sudden events that may be caused by instrumental errors or short-term anomalous celestial phenomena.

To recognize and remove or attenuate the noise in the high-frequency part, the wavelet coefficients on each scale have to be further thresholded to eliminate the noise further, and keep the most characteristic information. The key to this step of processing is to select the threshold value and threshold function suitable for the data in question.

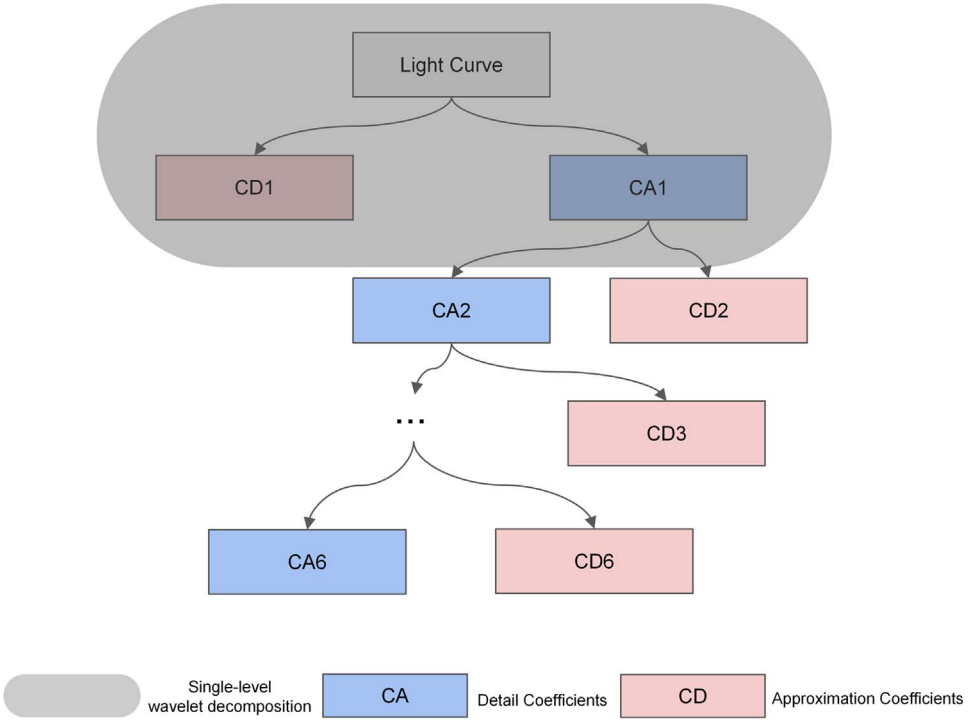

Figure 3. Wavelet transform 6-layer decomposition.

The median absolute deviation method was used to estimate the noise standard deviation of the data with the formula:

$$\sigma = (\text{median}(|x|))/0.6475. \tag{3}$$

Next, the soft threshold method is used, i.e., coefficients less than the threshold are set to 0, and coefficients greater than the threshold are subtracted from the threshold, with the formula:

$$\widehat{x_{i,j}} = \begin{cases} \text{sign}(x_{i,j} \cdot (|x_{i,j} - t|)), & |x_{i,j}| > t \\ 0, & |x_{i,j}| < t \end{cases}, \tag{4}$$

where x is the coefficient after performing the decomposition and t is the threshold value.

Here the threshold σ is thresholded using VisuShirk thresholding, which is a widely used thresholding method with the formula:

$$t = \sigma\sqrt{2\ln N}, \tag{5}$$

where t is the threshold, σ is the standard deviation calculated from the wavelet coefficients, and N is the number of sample points in a sample.

Finally, based on the thresholded wavelet coefficients, a discrete wavelet inverse transform (DWIT) is done to reconstruct the signal using the remaining frequency components, and the reconstructed light-variation curve is obtained as shown in Figure 4. The reconstructed signal is of high quality and well preserves the anomalous features of the original light-variation curves, such as peaks and mutations, which can reveal weak signals that were originally masked by noise, and may point to new astrophysical discoveries.

3.2. Optimization of LSTM Network Structure based on GRU

The structure of LSTM consists of the cell state, the current time step input, the previous time step hidden state, and three gates, which are forget gate, input gate, and output gate, as shown in Figure 5. The cell state is the core of LSTM, which can transfer information between time steps and be updated or retained by gate control.

The forgetting gate performs the first processing of the cell state, which determines the percentage of the cell state that is forgotten, and it uses the sigmoid function, which works by generating a value between 0 and 1. The more the value tends to 1, the more the information tends to be retained in its entirety, and the more the value tends to 0, the more the information tends to be completely forgotten, with the formula:

$$f_t = \sigma(W_f\ [h_{t-1}, x_t] + b_f), \tag{6}$$

where f_t is the output of the oblivion gate, W_f is the weight matrix of the oblivion gate, b_f is the bias vector of the oblivion gate, h_{t-1} is the hidden state of the previous time step, and x_t is the input of this current time step.

The input gate processes the cell state a second time, it takes the input information of the current time step into

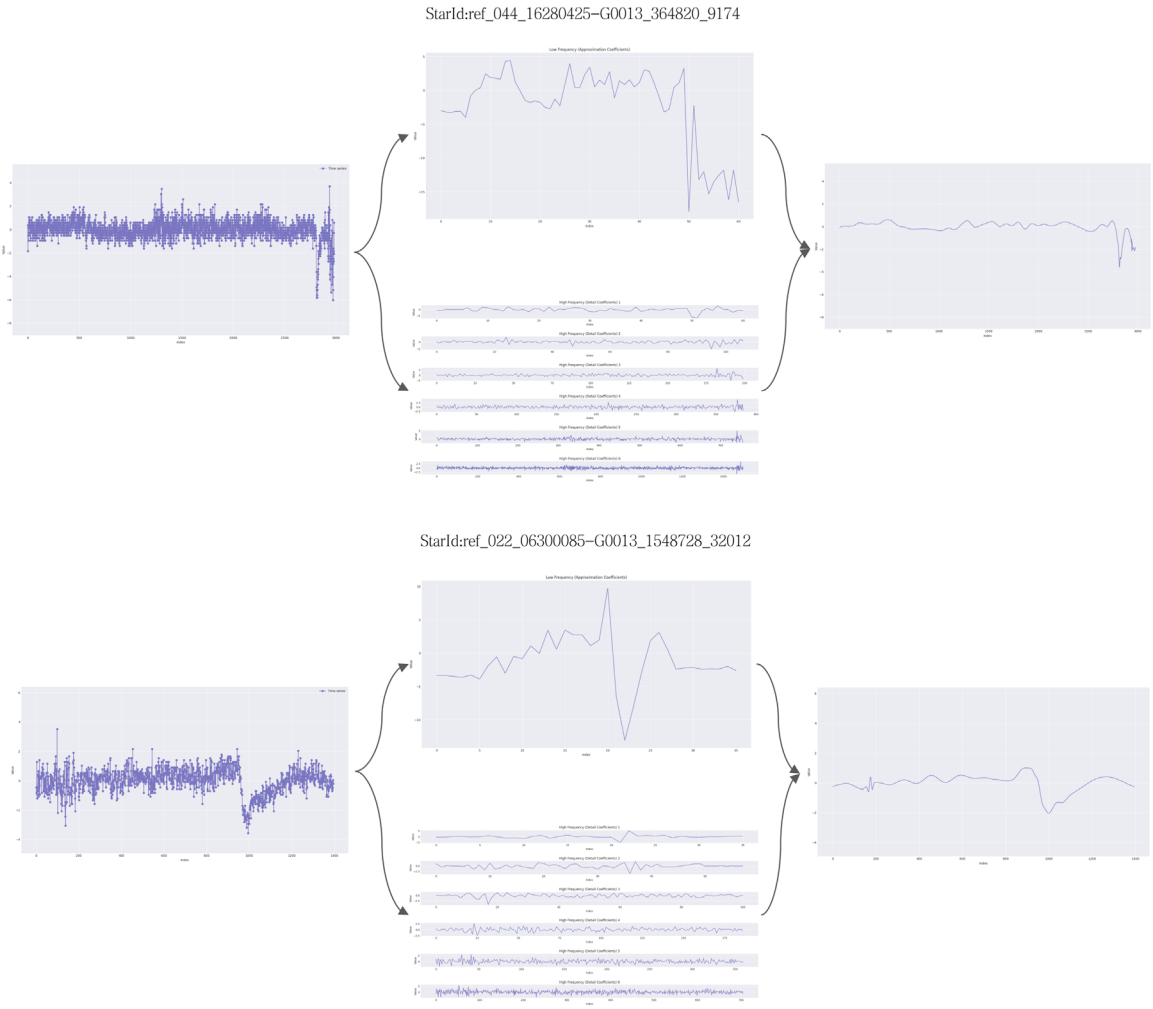

Figure 4. Wavelet transform process.

consideration, using the sigmoid function to generate a value between 0 and 1. If the value tends to 1, the current time step input information is added in full, and the more the value tends to 0, the less the input information of the current time step has been adopted. The tanh function is also used to output a candidate cell state that represents the input information. The formula for the input gate is:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{7}$$

$$\widehat{C_t} = \tanh(W_C[h_{t-1}, x_t] + b_c), \tag{8}$$

where i_t is the output of the input gate, W_i, W_C is the weight matrix of the input gate, b_i, b_c is the bias vector of the input gate, $\widehat{C_t}$ denotes the state of the cell after processing by the input gate, h_{t-1} is the hidden state of the previous time step, and x_t is the input of this current time step.

The output gate processes the cell state for the third time, which determines the part of the cell state information that is finally output. A value between 0 and 1 is generated using the sigmoid function.If the value tends to 1, the current time step output information is retained in its entirety, and the more the value tends to 0, the less the current time step out information is taken in. The role of the tanh function is to activate the cell state. The formula is:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \tag{9}$$

$$h_t = o_t * \tanh(C_t), \tag{10}$$

where o_t is the output of the current output gate, h_t is the hidden state of this time step, W_o is the weight matrix of the output gate, b_o is the bias vector of the output gate, C_t is the cell state of this time step, and $*$ denotes the Hadamard product.

GRU has faster speed and accuracy in training light curve data, it is a variant of LSTM network which can solve the problem of long term dependency, i.e., using the past information to influence the future output, there are dependencies in the light curve time series, GRU is more adaptable to

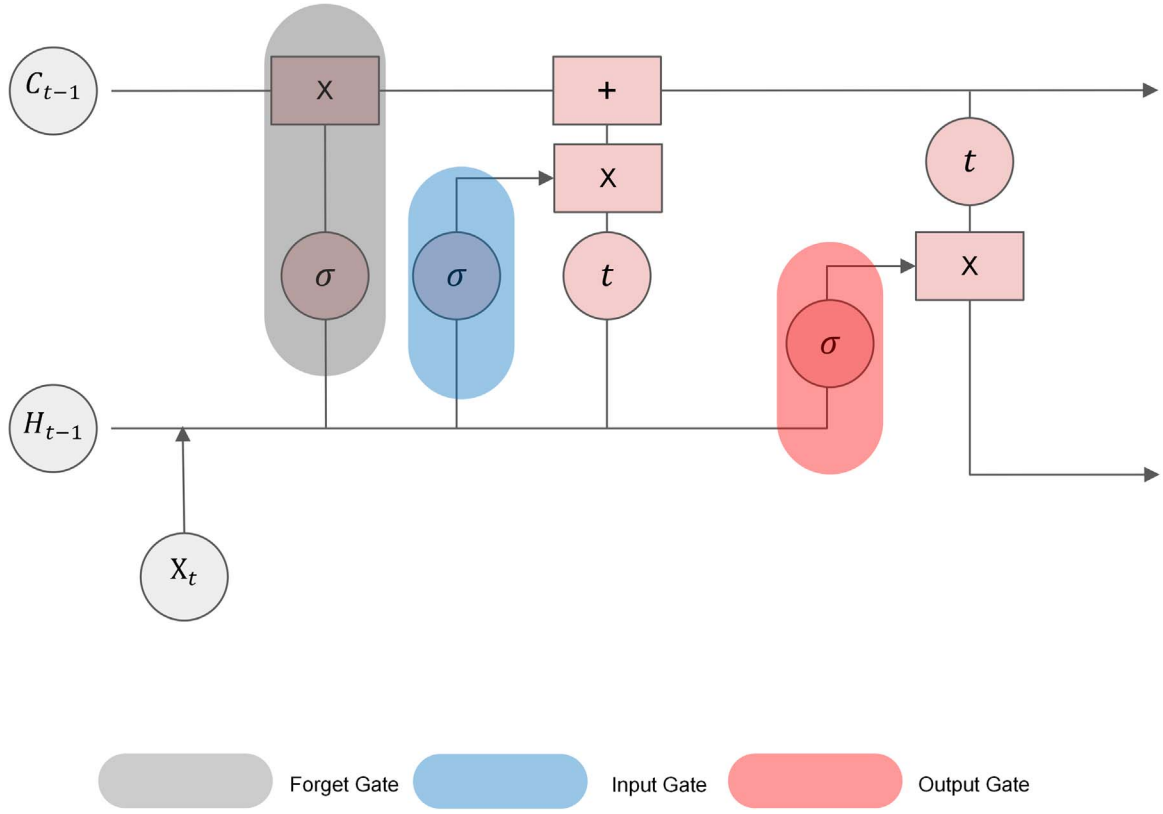

Figure 5. LSTM network structure.

detect anomalies of light curve, and at the same time improve the efficiency of the network, the design of GRU helps to reduce the problem of vanishing gradient. The design of GRU helps to reduce the gradient vanishing problem, which is more favorable for model training when dealing with long data series, and this feature is also more applicable to the light curves with long time series. The GRU network consists of a hidden state, the input of the current time step, and two gates, which are the update gate and the reset gate. The hidden state is the most important part of the GRU network, which can pass the information from the first layer to the last layer and decide whether to update or retain through the gates, the structure is shown in Figure 6.

The reset gate does the first processing of the hidden state, which is able to selectively reset, using a sigmoid function to output a value between 0 and 1, indicating that the information will tend to be not reset or completely reset. The formula is:

$$r_t = \sigma(W_r[h_{t-1}, x_t] + b_r), \tag{11}$$

where r_t is the output of the reset gate, W_r is the weight matrix of the reset gate, b_r is the bias vector of the reset gate, h_{t-1} is the hidden state of the previous time step, and x_t is the input of this current time step.

The update gate does a second processing of the hidden state, it can selectively update from the hidden state, using the sigmoid function to output a value between 0 and 1, depending on the size of the value to decide to tend to no update or all update. The formula is:

$$z_t = \sigma(W_z[h_{t-1}, x_t] + b_z), \tag{12}$$

where z_t is the output of the update gate, W_z is the weight matrix of the update gate, b_z is the bias vector of the update gate, h_{t-1} is the hidden state of the previous time step, and x_t is the input of this current time step.

3.3. Attention Mechanism—Time Series Weight Assignment

Attention mechanism is a technique used to improve the generalization and robustness of neural networks and the efficiency of network performance, the model can get a better training effect and improve the accuracy of the model by paying extra attention to the key or relevant parts when processing sequence data, so that the key part of the light change curve that determines the anomaly will be given higher weight and get good results, which can provide the astronomical observations and research to provide valuable information and guidance. The most central formula of the

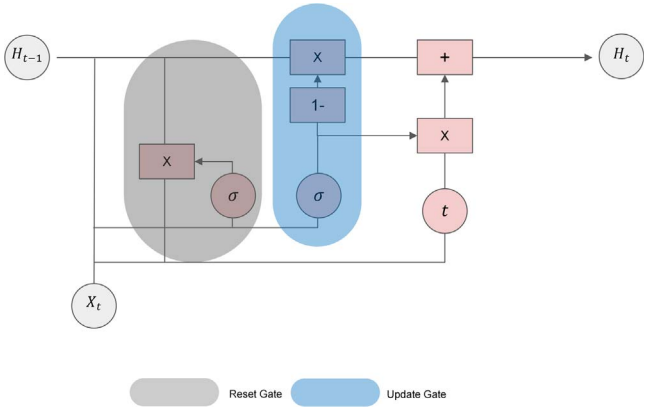

Figure 6. GRU network structure.

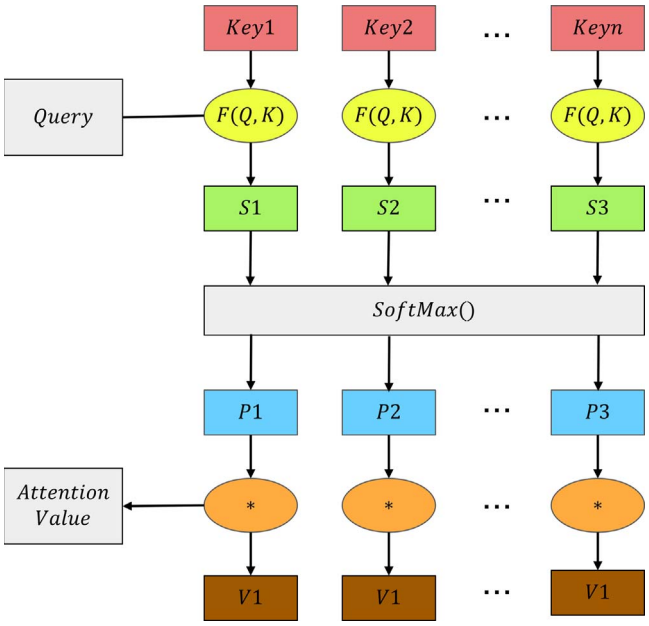

Figure 7. Structure of Attention Mechanism.

attention mechanism Vaswani et al. (2017) is:

$$\text{Attention}\,(Q,\,K,\,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) * V, \qquad (13)$$

where, Q, K and V denote the query (Query), key (Key), and value (Value) respectively, which are obtained from the hidden state vectors of the input or output sequences after some transformations. Q and the dot product of K is the degree of similarity between these two parameters, and dividing it by d_k is for scaling and stabilizing the gradient, and the function normalizes the similarity to a probability distribution, which is used as the weight of each V . Finally, the weighted V vectors are summed to get the context vector as the output of the *Attention*, as shown in Figure 7.

3.4. DW-GRU-Attention Model—Improved GRU Structure Based on Attention and Wavelet

3.4.1. Overall Structure

The model uses the wavelet change for feature extraction and signal denoising of the light-change curve data, and then uses the gated recurrent unit network and the attention mechanism to process the output of the wavelet transform, and finally realizes the prediction of the light-change curve anomalies. Wavelet transform can remove the noise of light curve while preserving the original data features, and reveal the weak abnormal signals that have been covered up originally. The gated recurrent unit network can utilize the past information to influence the output in the future, and keep the information in the beginning to the end, and at the same time, solve the problem of long-term dependence on the data, which can play a very good role in the prediction of the light curve time-series data, and at the same time, attention mechanism can be used to find the part of the time series that plays a key role in detecting anomalies by giving different time step weights.

3.4.2. Model Shape Design and Structure Design

In this study, the advantages of the three techniques are synthesized to design the neural network structure, and the shape of the model hierarchical design is shown in Figure 8.

The training process of the algorithm is shown in Table 3.

The input layer converts the data of the light curves into vector representations as input sequences.

The wavelet transform layer decomposes the original light curves into components of different scales according to frequency, thresholding the slow trend of the signal mainly described by the low-frequency part as well as the fast variations and details of the signal captured in the high-frequency part before inverting them.

Bidirectional GRU is used to encode the input sequence to obtain the hidden state vector of the time step. The bidirectional GRU enhances the expressive power of the model by simultaneously considering the contextual information before and after.

The output of the GRU layer is weighted and averaged using the attention mechanism to obtain a global context vector. The attention mechanism allows the model to focus on the most important parts of the input sequence, improving the accuracy of the model.

Finally, the context vector is mapped to a scalar representing the probability of an anomaly warning using a fully connected layer and an activation function. The activation function uses softmax to determine whether or not to signal an alert based on a preset threshold.

The network structure is shown in Table 4, the input layer of the model is the light-variation curve data in the format of n^*1,

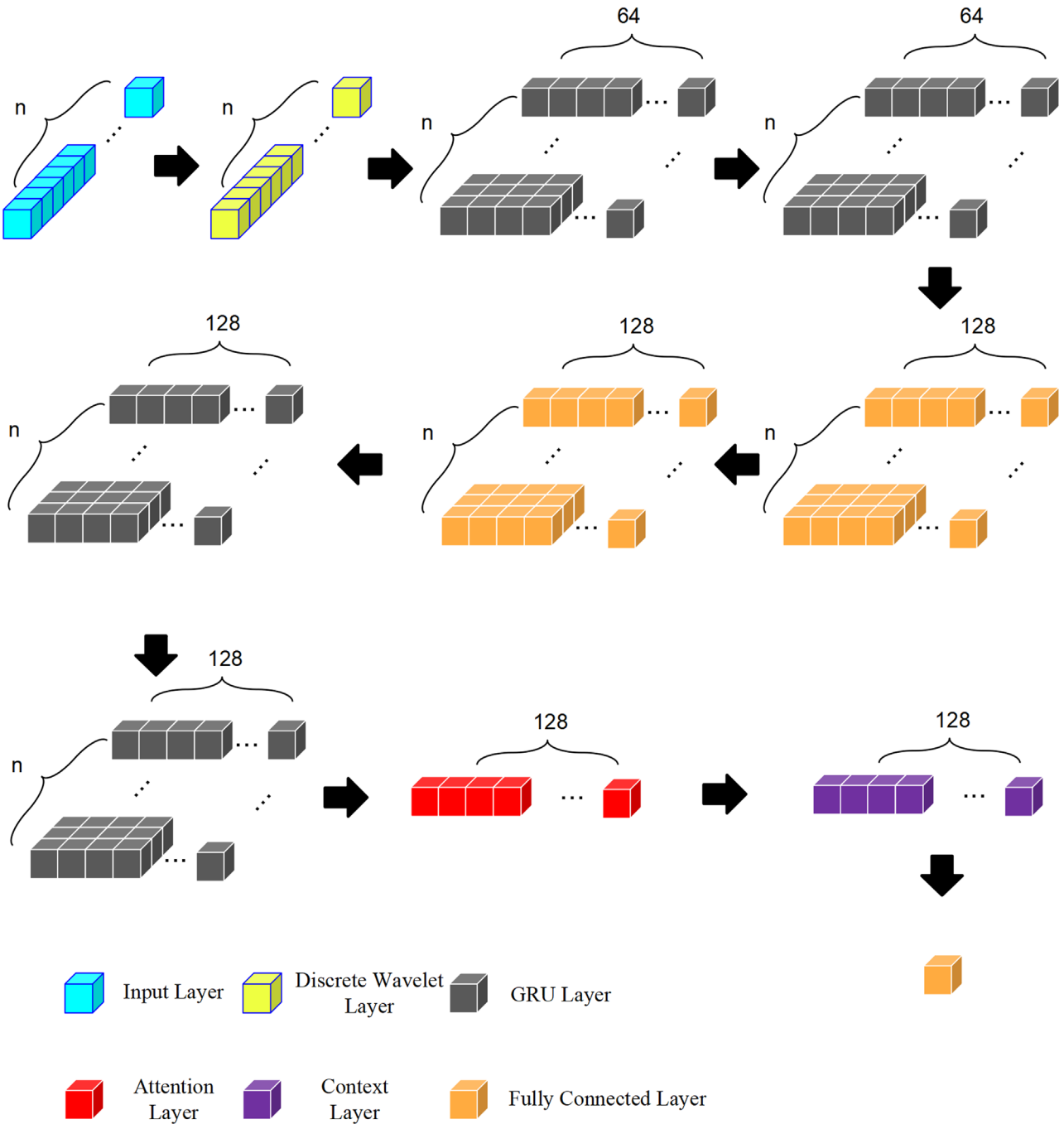

Figure 8. DW-GRU-Attention model layered design shape.

followed by two GRU layers, here the hidden layer dimension is 64, so the output dimension is n^*64, followed by a fully connected layer that maps the data from the output dimensions of the GRU layer to larger dimensions, and after mapping is complete, the Dropout layer is used to prevent overfitting. Continue to add two GRU layer, the output is followed by the attention layer, the output dimension is 1^*128, the dimensions are the batch size and sequence length, followed by the attention weight and the output to calculate the weighted sum, the dimensions are the batch size and the hidden layer dimensions, and finally, the context vector is input into the fully connected layer to get the final output, the dimensions of 1, that is, to get the next prediction, the next prediction. Here also the fully connected output dimension can be modified and

Table 3
Algorithmic Training Process

Algorithm The Training Process of DW-GRU-Attention
Input:
Number of training sets
Number of test sets
Light Curve Dataset
training times
Output:
predicted value
1:The input layer converts the data from the light curves into a vector representation as an input sequence.
2:Decompose the original light curve into components of different scales according to frequency.
3:Bidirectional GRU is used to encode the input sequence to obtain the hidden state vector of the time step.
4:The output of the GRU layer is weighted and averaged using the attention mechanism to obtain a global context vector.
5:Mapping a context vector to a scalar using a fully connected layer and an activation function.

Table 4
DW- GRU-Attention Model Structure

Layer No.	Layer Name	Activation Function	Input Dimension	Output Dimension
1	Input Layer	⋯	⋯	(n, 1)
2	Wavelet Layer	⋯	(n, 1)	(n, 1)
3	GRU Layer 1	tanh	(n, 1)	(n, 64)
4	GRU Layer 2	tanh	(n, 64)	(n, 64)
5	Full Connected Layer	ReLU	(n, 64)	(n, 128)
6	Dropout Layer	⋯	(n, 128)	(n, 128)
7	GRU Layer 3	tanh	(n, 128)	(n, 128)
8	GRU Layer 4	tanh	(n, 128)	(n, 128)
9	Attention Layer	Softmax	(n, 128)	(n, 128)
10	Context Layer	⋯	(n, 128)	(n, 128)
11	Full Connected Layer	ReLU	(n, 128)	(1, 1)

multiple values can be predicted, the structural design of DW-GRU-Attention is shown in Figure 9.

4. Analysis and Comparison of Experimental Results

4.1. Evaluation Criteria

Here the F1 parameter is used as a metric to assess the completeness of the warning, by manually dividing the anomaly intervals and evaluating and comparing the anomaly detection results with the anomaly intervals derived from the model prediction, and the same anomaly intervals will be used as a criterion for all the models. The F1 parameter is a metric used to assess the performance of the classification problem, which combines the performance of the two aspects of the precision rate and the recall rate. The two metrics, precision rate and recall rate, are used to evaluate the performance of classification problems, and they are both based on the four values of the confusion matrix: true positive, true negative, false positive, and false negative, precision is the proportion of correctly categorized positive instances to all predicted positive instances, and recall is the proportion of correctly categorized positive instances to all true positive instances. The F1 formula is:

$$\mathrm{F1} = \frac{2 * \mathrm{precision} * \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}. \tag{14}$$

The formulas for precision and recall are:

$$\mathrm{precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{15}$$

$$\mathrm{recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{16}$$

where TP denotes the true example, FP denotes the false positive example, and FN denotes the false negative example.

4.2. Experiments with Simple LSTM Models

The method used by Zhang & Zou (2018) is LSTM without wavelet transform for anomaly detection of time series, first, the LSTM model is trained using light-variable curves without wavelet transform, the learning rate used is 0.00001, the number of training times is 50, the training set and test set are thirty-five percent before and after the target star time series, the number of hidden layers is 64, and the size of sliding window is 2. The target starIds are ref_033_16810765-G0013_482792_32012 and ref_044_16280425-G0013_482792_32012 respectively, and both have anomalies.

As the training proceeds, the accuracy of the model increases, as in Figure 10, the model can determine whether there is an anomaly in the light curve, but the F1 scores are low, 0.138 and 0.329, respectively, and the anomalies derived from the model do not fully cover all the real anomalies, i.e., the completeness of the model prediction needs to be improved, and the prediction results are visualized in Figure 11, where the successful detection of anomalies is shown on the upper side, and the confusion matrix is shown on the lower side of the image. The top side of the image shows the successfully detected anomalies, and the bottom side shows the confusion matrix.

4.3. Experiments with Simple GRU Models

The method used by Rui-Qing Yan [14] is GRU without wavelet transform for anomaly detection in time series, first the GRU model is trained using light change curve without wavelet transform, the learning rate used is 0.00001, the number of training times is 50, the training set and test set are 35 percent before and after the time series of the target star respectively, the number of hidden layers is 64. The sliding window size is 2. The model f1 values are shown in Figure 12, which are 0.258 and 0.359, respectively, and the results in f1 values are improved by

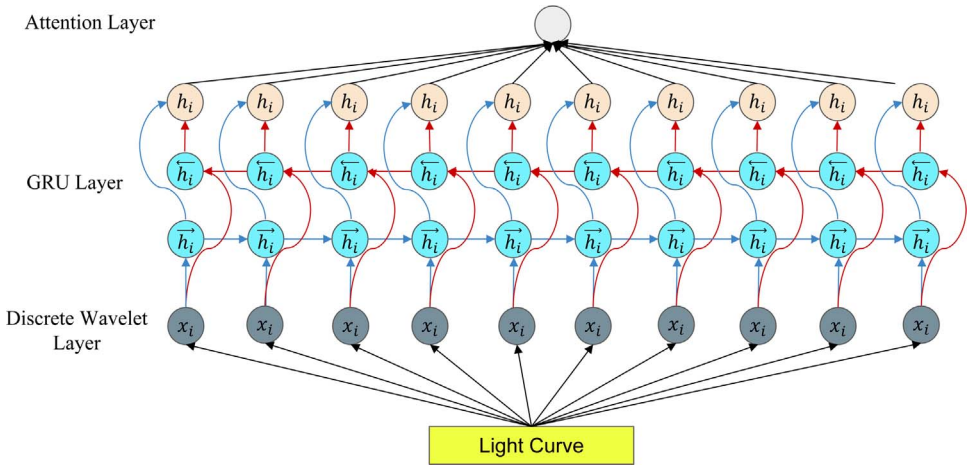

Figure 9. Design of DW-GRU-Attention network structure.

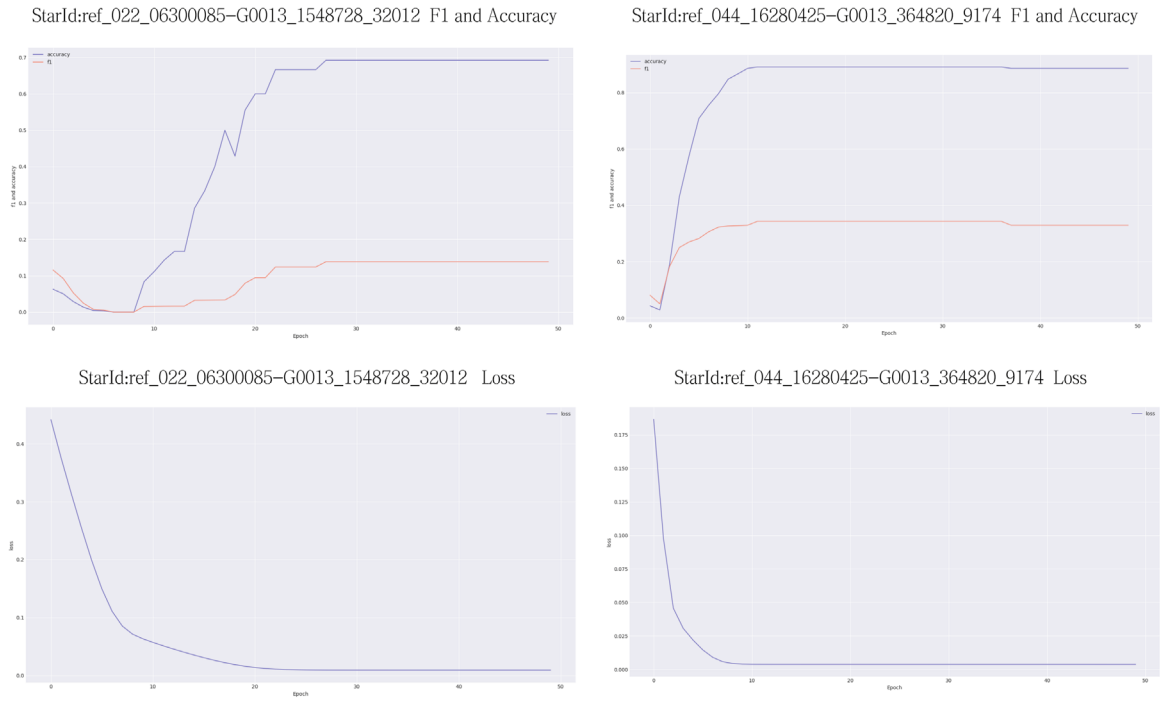

Figure 10. Experimental F1, Accuracy, Loss of LSTM model.

0.12 and 0.03 on the basis of LSTM, which achieves a certain optimization effect, and the model visualization results and confusion matrix are shown in Figure 13.

4.4. Experiments with GRU Models Using Wavelet Transforms

From Figure 14, it can be seen that the light-variation curves have better expressiveness on the GRU model after wavelet transform processing, and the wavelet transform reduces the noise while retaining the original features, the StarId of ref_033_16810765-G0013_482792_32012 and ref_044_16280425-G0013_364820_9174. The final F1 of the two stars is stabilized at around 0.812 and 0.760, which is improved by 0.554 and 0.401 in f1 compared with the GRU model method without wavelet transform, which improves the complete identification of the anomalies of the light-variation curves while doing the determination of whether there are any anomalies in the light-variation curves, so that more anomalies can be identified, and at the same time the number of training

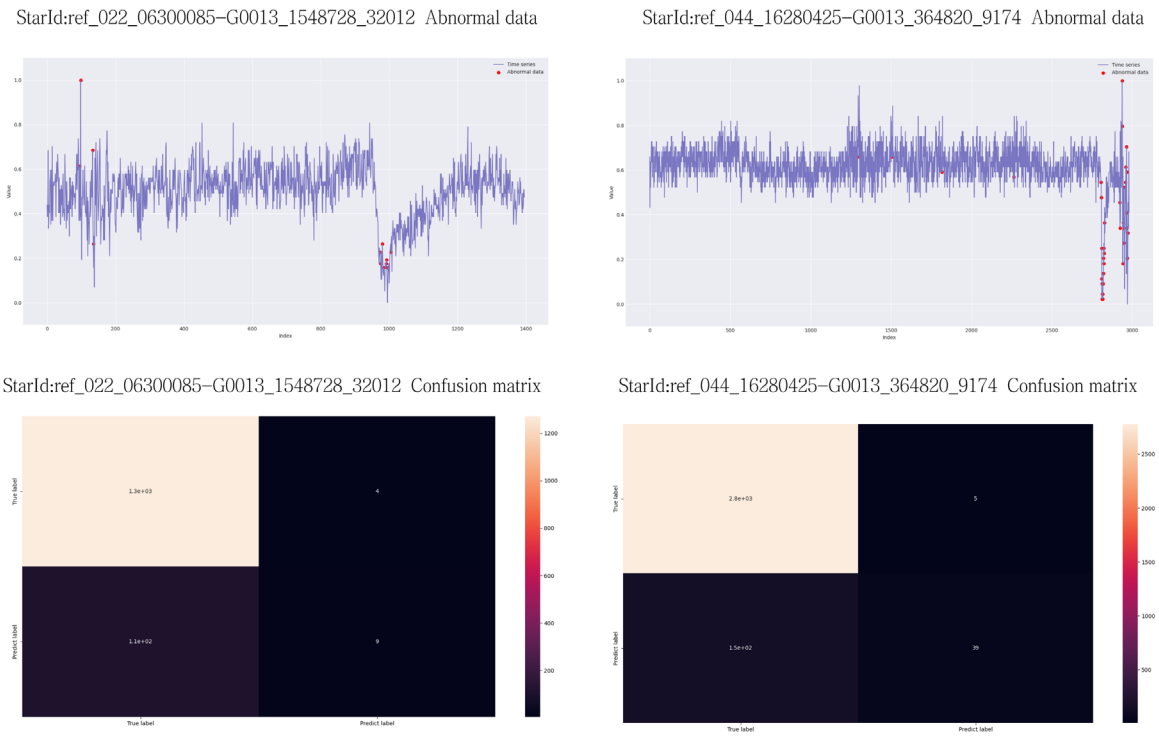

Figure 11. Experimental prediction results and confusion matrix of LSTM model.

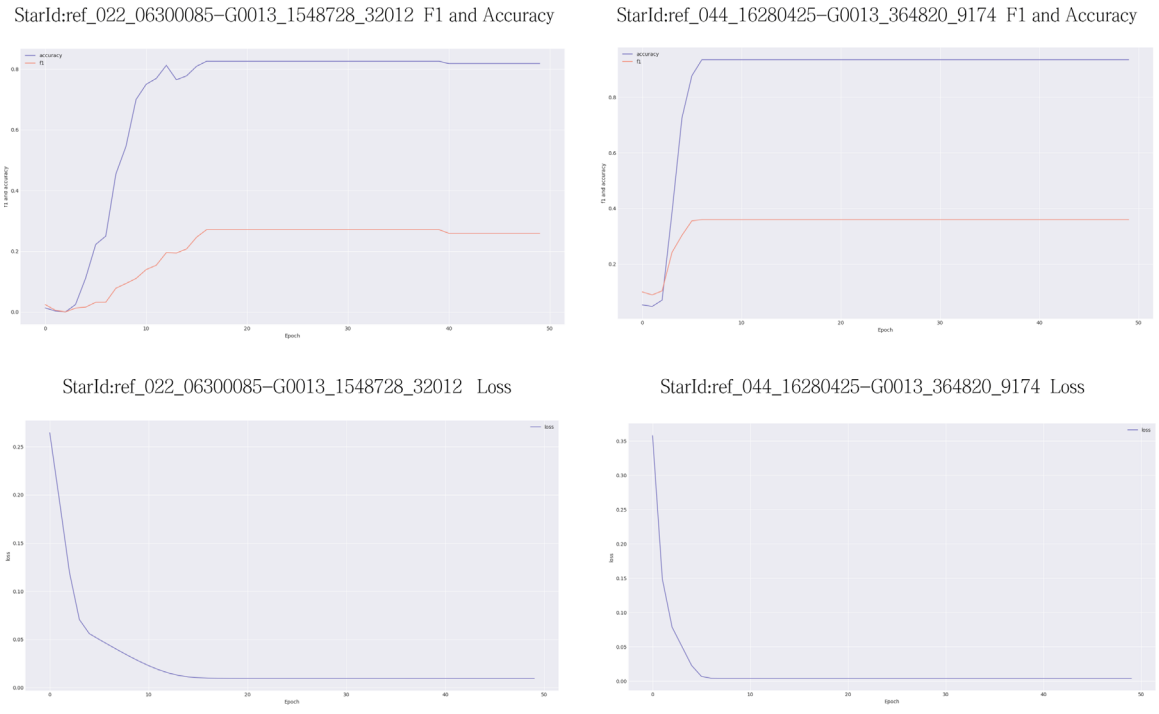

Figure 12. Experimental F1, Accuracy, Loss of GRU model.

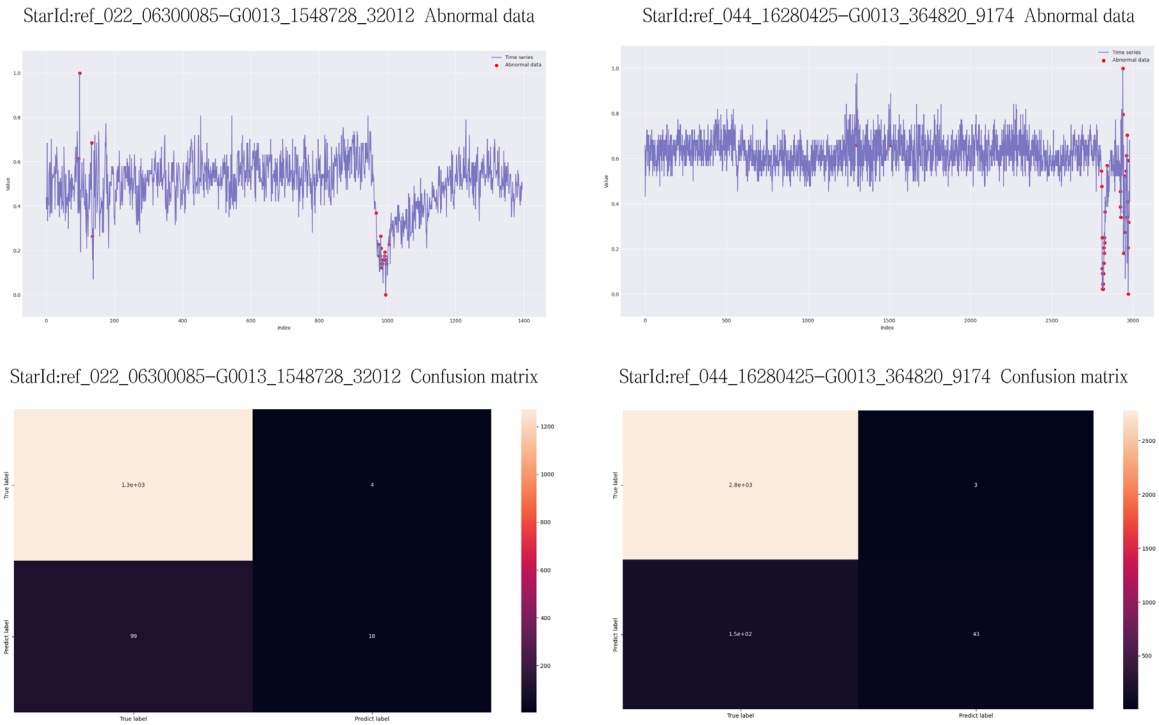

Figure 13. Experimental prediction results and confusion matrix of GRU model.

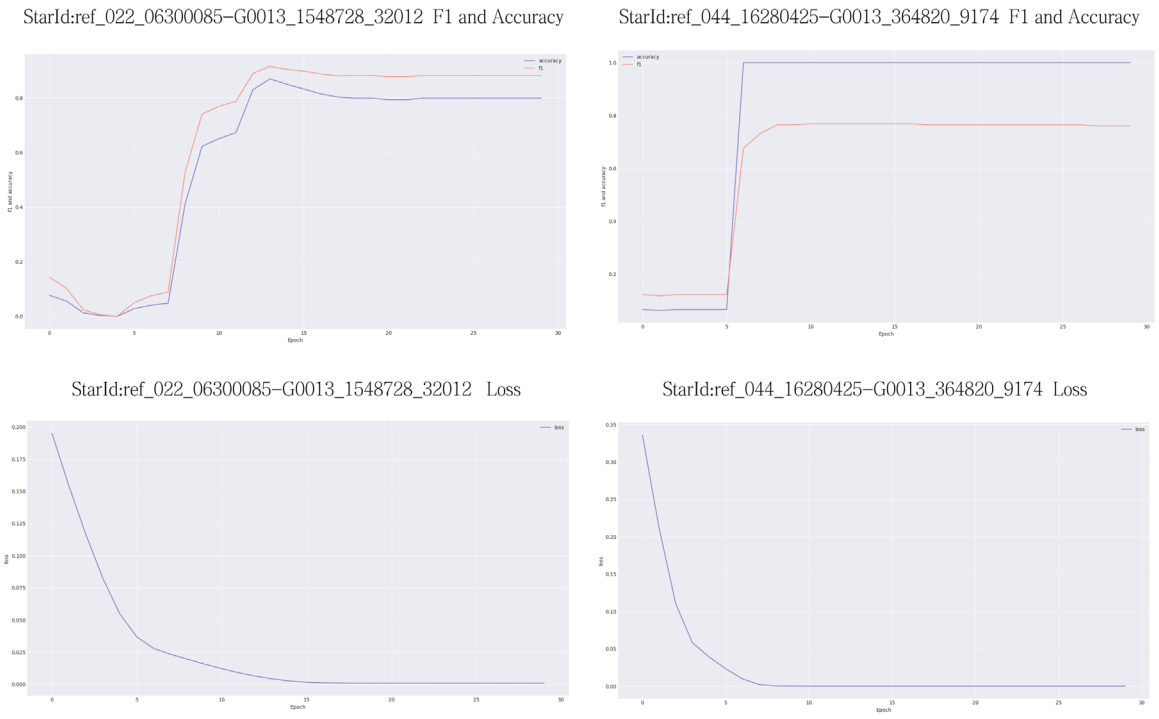

Figure 14. Experimental F1, Accuracy, Loss of GRU model using wavelet transform.

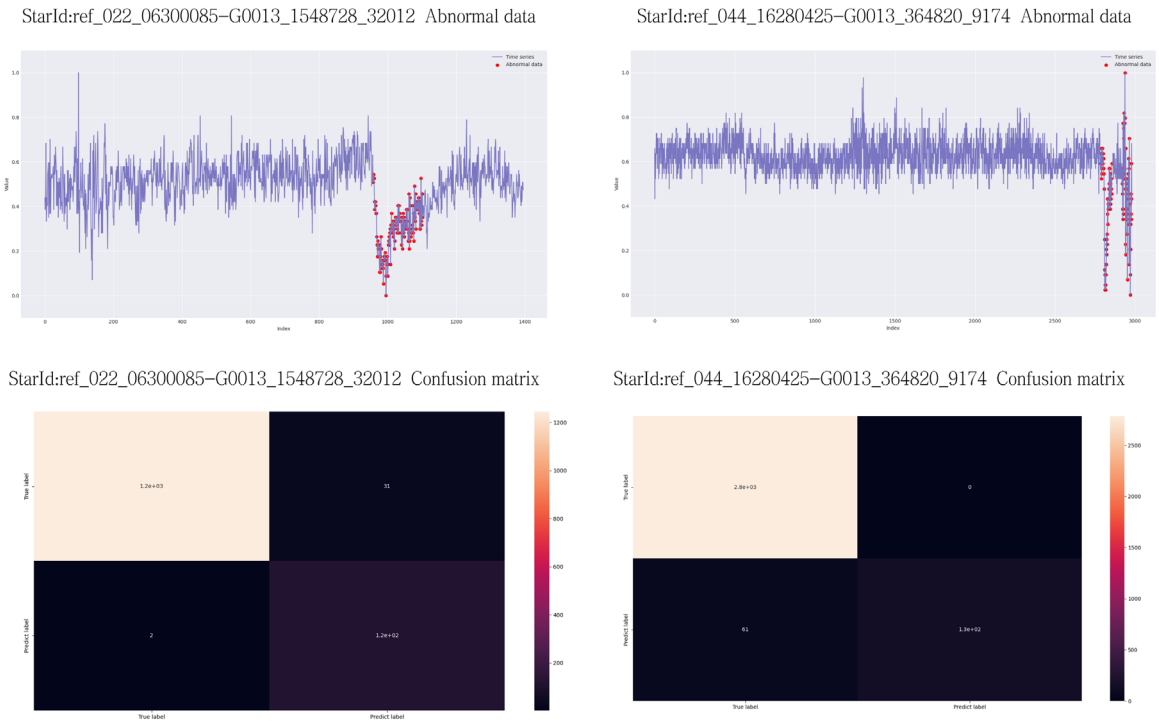

Figure 15. Experimental prediction results and confusion matrix of GRU model using wavelet transform.

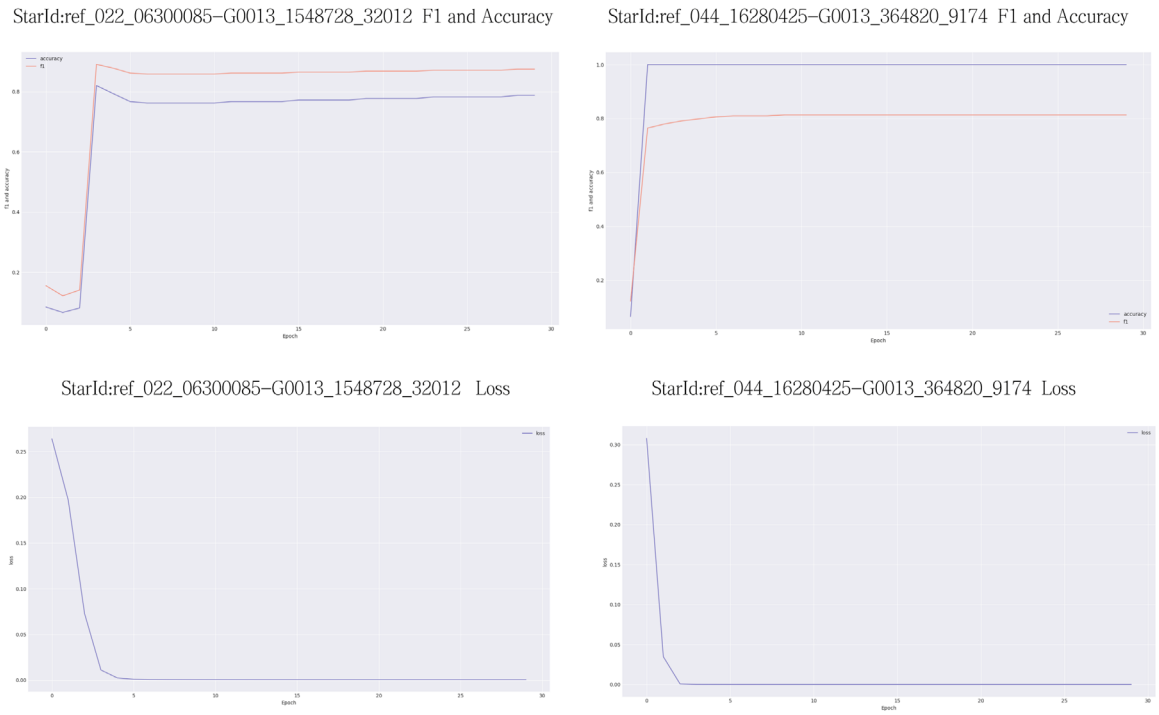

Figure 16. Experimental F1, Accuracy, Loss of DW-GRU-Attention model using wavelet transform.

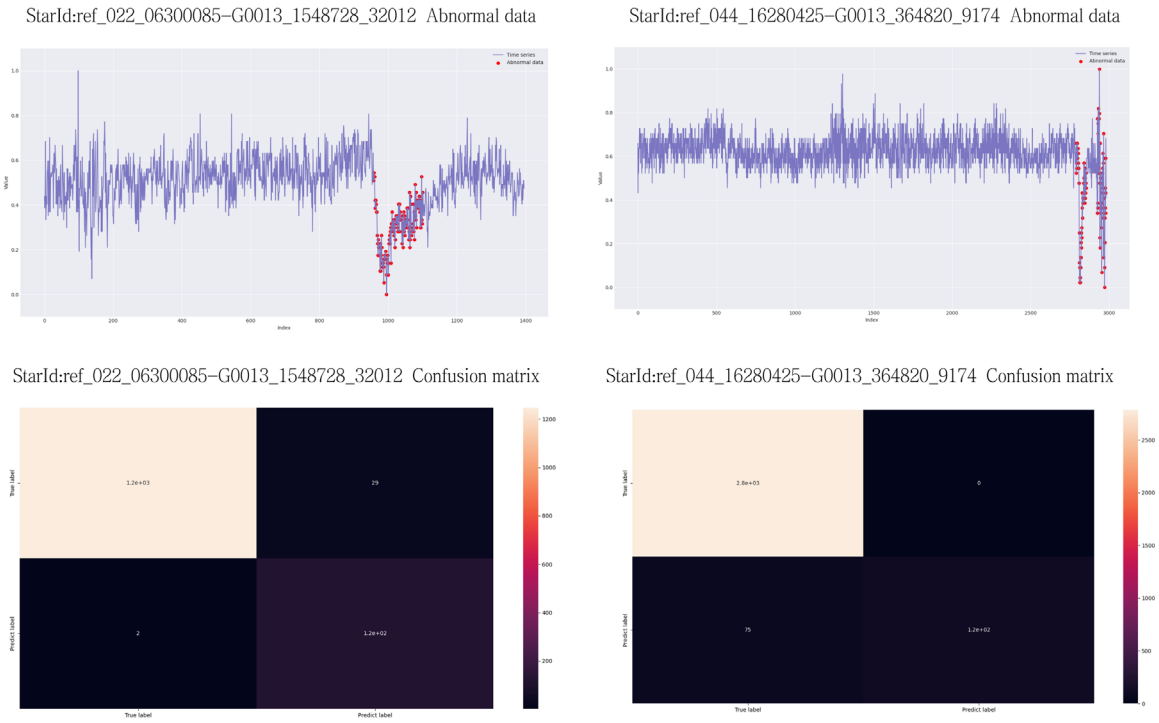

Figure 17. Experimental prediction results and confusion matrix of DW-GRU-Attention model using wavelet transform.

Table 5
Comparison of Warning Effects of Different Models

StarId	Model No.	Model Name	F1 Score	Stable Convergence Training Times
32012	1	LSTM	0.138	27
32012	2	GRU	0.258	18
32012	3	DW-GRU	0.812	17
32012	4	DW-GRU-Attention	0.874	5
9174	1	LSTM	0.329	11
9174	2	GRU	0.359	7
9174	3	DW-GRU	0.760	6
9174	4	DW-GRU-Attention	0.813	2

times is reduced, and the training efficiency is effectively improved, and the prediction results are visualized in Figure 15.

4.5. Experiments with the DW-GRU-Attention Model Using the Wavelet Transform

This experiment then adopts attention to improve the GRU, which is intended to improve the efficiency of the neural network, and introduces the attention mechanism to construct the GRU-Attention neural network as described in the previous section. This obtains further improvement in the completeness of the light-change curve warning, the results are shown in Figure 16, and the final f1 is stabilized at 0.874 and 0.813, which is 0.062 and 0.053, respectively, and improves further in terms of efficiency, while the training efficiency is effectively improved, and very good convergence is achieved at three times of training. It has excellent performance in light-change curve anomaly detection, and the final anomaly detection result and confusion matrix, as shown in Figure 17.

4.6. Comparison of Results

In this paper, the DW-GRU-Attention light curve early warning model is compared with the LSTM method of Zhang & Zou (2018), and the results are shown in Table 4. In the data set, the f1 values of the two stars ref_033_16810765-G0013_482792_32012 and ref_044_16280425-G0013_364820_ The f1 values of the anomaly detection results for 9174 are 87.4% and 81.3%, which are improved by 73.6% and 48.4%, respectively.

Comparing with the GRU method of Yan et al. (2020), the results are shown in Table 5, where the f1 values of the anomaly detection results of the two stars ref_033_16810765-G0013_482792_32012 and ref_044_16280425-G0013_364820_9174 in the data set are improved by 61.6% and 45.4%.

Previous paper work mainly detects whether the object is anomalous or not, and fails to detect all the anomalous time nodes, the method proposed in this paper can cover most of the anomalous time nodes, and the addition of the attention

mechanism makes the key part of the light-variation curves determining the anomalies to be given a higher weight, and in the actual anomalies detection, the two stars are detected with 98.2% and 68.5% of anomalies, respectively, while having less training time.

5. Summary

This paper summarizes and discusses on the basis of previous research. Based on the shortcomings of previous researchers in the light-variation curve anomaly detection in terms of poor data feature extraction, low prediction accuracy, and low efficiency of the prediction model. We propose the DW-GRU-Attention light-variation curve early warning model. The signal noise of the light curve is complex, and the subtle information related to the anomalies is easily hidden in the noise. In this paper, the wavelet transform is used to decompose the light curve time series data into six layers of features, so that the light curve features can be retained under the premise of removing the signal noise as much as possible. Meanwhile, the time series of light-variable curve is long, the introduction of gated recurrent unit network greatly improves the efficiency of the model, and this paper incorporates the attention mechanism to find out the key parts of the light curve that determine the anomalies, and these parts are assigned higher weights. Finally, the experimental results were evaluated by f1 value, accuracy, confusion matrix, and visual anomaly analysis. By comparing with the GRU, LSTM, and DW-GRU methods, the f1 values on the stars were improved by 61%, 53.5%, and 5.75% on average, which indicates that the method in this paper achieves a better result and possesses a higher efficiency, and proves that the model possesses an excellent performance, and at the same time, and there is room for improvement in the effective feature retention of light-variation curve processing and real-time light-variation curve warning, need to be studied further. This model opens up a new way, which thinking for the future research of light curves, i.e., applying the weights of the attention mechanism, to the detection of astronomical big data, which can capture the features, that are difficult to be captured by the traditional statistical methods in the past more conveniently, and at the same time complement each other with the wavelet transform, which can provide valuable help for the research of astronomers.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant id: 2022YFF0711500), the National Natural Science Foundation of China (grant id: 11803022 and grant id: 1227307712273077).

ORCID iDs

Hao Li https://orcid.org/0009-0004-2623-286X

References

Althukair, A., & Tsiklauri, D. 2023, RAA, 23, 085017
Bi, J., Feng, T., & Yuan, H. 2018, Comput. Ind., 97, 76
Boone, K. 2021, AJ, 162, 275
Bowles, M., Bromley, M., Allen, M., & Scaife, A. 2021, arXiv:2111.04742
Breton, S. N., Santos, A. R., Bugnet, L., et al. 2021, A&A, 647, A125
Burhanudin, U., Maund, J., Killestein, T., et al. 2021, MNRAS, 505, 4345
Chakraborty, N. 2019, Preprints, 2019, 2019070241
Deb, S., & Singh, H. P. 2009, A&A, 507, 1729
Feng, T., Du, Z., Sun, Y., et al. 2017, in 2017 IEEE Int. Congress on Big Data (Bigdata Congress) *(Honolulu, HI, 25-30 June 2017)* (Piscataway, NJ: IEEE), 224
Huang, T. 2019, Master's thesis, University of Science and Technology of China
Kalaee, M., & Hasanzadeh, A. 2019, NewA, 70, 57
Lu, C., Peng, L., Bi, J., & Yuan, H. 2018, in 2018 5th IEEE Int. Conf. Cloud Computing and Intelligence Systems (CCIS) *(Nanjing, China, 23-25 November 2018)* (Piscataway, NJ: IEEE), 117
Lu, L. 2022, Master's thesis, Yunnan Normal University
Ma, J., Ma, S., Ziming, Z., & Li, Y. 2022, PrA, 4, 575–589
Sasal, L., Chakraborty, T., & Hadid, A. 2022, in 2022 21st IEEE Int. Conf. Mach. Learn. Appl. (ICMLA) *(Nassau, 12-14 December 2022)* (Piscataway, NJ: IEEE), 671
Van Doorsselaere, T., Shariati, H., & Debosscher, J. 2017, ApJS, 232, 26
Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, Advances in Neural Information Processing Systems (Curran Associates, Inc.), 30
Vida, K., & Roettenbacher, R. M. 2018, A&A, 616, A163
Xu, L., Yu, X., & Yan, Y. 2018, E-science Technology and Application, 3, 49
Xu, P., Cao, J., Chen, W., et al. 2022, Beijing Da Xue Xue Bao, 58, 808
Yan, R.-Q., Liu, W., Zhu, M., et al. 2020, RAA, 20, 007
Yu, C., Li, K., Zhang, Y., et al. 2021, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11, e1425
Zhang, R., & Zou, Q. 2018, J. Phys.: Conf. Ser., 1061, 012012